# Lec 9 MATH 343 Inference for linear regressions

$\mathbb{D} = \langle X, \vec{y} \rangle$, $X$ has an intercept col so $X \in \mathbb{R}^{n \times (p+1)}$, $\vec{y} \in \mathbb{R}^n$

$\mathcal{H} = \{\vec{w}\vec{x} : \vec{w} \in \mathbb{R}^{p+1}\}$ Assume $\mathcal{A}$: minimize SSE returns $g(\vec{x}) = \vec{x}\vec{b}$

$\Rightarrow \vec{b} = (X^\top X)^{-1} X^\top \vec{y}$, $\hat{\vec{y}} = X\vec{b} = H\vec{y}$, Also: $\vec{e} := \vec{y} - \hat{\vec{y}}$
$\Rightarrow \vec{y} = X\vec{b} + \vec{e}$

Is $g(\vec{x}) = h^*(\vec{x})$? No... there is estimation error.

In this topic, we assume goal is to find $h^*(\vec{x})$, not $f(\vec{x})$.

Recall $h^*(\vec{x}) = \vec{x}\vec{\beta} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$. $\dim[\vec{\beta}] = p+1$, thus the entries of $\vec{\beta}$ are the parameters of interest.

Thus $\vec{b}$ is a point estimate for $\vec{\beta}$. How do we get frequentist CI's, hypothesis tests, Bayesian CR's, Bayesian tests?

We need to assume r.v.'s somewhere!

$$y_i = h^*(\vec{x}_i) + \epsilon_i \Rightarrow y_i = \vec{x}_{i\cdot}\vec{\beta} + \epsilon_i \Rightarrow \vec{y} = X\vec{\beta} + \vec{\epsilon}$$

= misspecification error + ignorance error for the $i$th subject

Consider $\epsilon_i = u_1 + u_2 + \ldots$, the sum of many, many unknown natures, likely independent of each other

If so, $\epsilon_i$ is likely a realization from a... normal distr! (Assum #1)

Which mean and variance? Since it's an error distr, mean zero for all $i$! (Assum #2)

Let variance be $\sigma_i^2$. Let's assume for now, that this variance is shared across all units i.e. $\sigma_i^2 = \sigma^2 \; \forall i$

Homoskedasticity assumption (Assum #3).

Assuming these three $\Rightarrow$ $\mathcal{E}_1, \ldots, \mathcal{E}_n \overset{iid}{\sim} N(0, \sigma^2)$ $\quad$ $\underset{\text{r.v.}}{\underset{\uparrow}{\mathcal{E}_i}}$ vs $\underset{\text{realization}}{\underset{\uparrow}{\epsilon_i}}$

$$\Rightarrow \vec{\mathcal{E}} \sim N_n\left(\vec{0}_{n}, \sigma^2 I_n\right) \Rightarrow \underset{\substack{\uparrow \\ \text{the response is now a r.v.} \\ \text{since } \vec{\mathcal{E}} \text{ is a r.v.}}}{\vec{Y}} = X\vec{\beta} + \vec{\mathcal{E}} \Rightarrow \vec{Y} \sim N_n\left(X\beta, \sigma^2 I_n\right)$$

What about $X$? For the purpose of this class, let $X$ be a fixed set of data. Otherwise things get really complicated!

$\vec{\beta}$ is the fixed parameters. So we assumed randomness in $\vec{\mathcal{E}}$, we assumed <u>no</u> randomness in $X$ and $\vec{\beta}$ are the fixed parameters thus $\vec{Y}$ is random only through $\vec{\mathcal{E}}$.

---

Under this setup, $\vec{b}$ is a point estimate for $\vec{\beta}$.

$\vec{b} = (X^T X)^{-1} X^T \vec{y} \Rightarrow \vec{B} = (X^T X)^{-1} X^T \vec{Y}$ the estimator for $\beta$

~~$\hat{\sigma}^2_{MLE} = \frac{1}{n} \sum e_i^2 \Rightarrow \hat{\sigma}^{2\,MLE} = \frac{1}{n} \sum E_i^2$, the estimator for $\sigma^2$, where $E_i$ are the r.v.'s whose realizations are the $e_i$'s.~~

We need the distribution of $\vec{B}$ if we're going to have inference for $\vec{\beta}$.

Recall $\vec{Z} \sim N_n(0, \sigma^2 I_n)$, $\vec{\mu} \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n} \Rightarrow \vec{\mu} + A\vec{Z} \sim N_m(\vec{\mu}, \sigma^2 A A^T)$

$$\vec{B} = (X^T X)^{-1} X^T \vec{Y} = (X^T X)^{-1} X^T (X\vec{\beta} + \vec{\mathcal{E}}) = (X^T X)^{-1} X^T X \vec{\beta} + (X^T X)^{-1} X^T \vec{\mathcal{E}}$$

$$= \vec{\beta} + (X^T X)^{-1} X^T \vec{\mathcal{E}} \sim N_{p+1}\left(\vec{\beta}, \sigma^2 (X^T X)^{-1} X^T \left((X^T X)^{-1} X^T\right)^T\right)$$

$$= N_{p+1}\left(\vec{\beta}, \sigma^2 (X^T X)^{-1} X^T X \left((X^T X)^{-1}\right)^T\right) = N_{p+1}\left(\vec{\beta}, \sigma^2 \left((X^T X)^T\right)^{-1}\right) = N_{p+1}\left(\vec{\beta}, \sigma^2 (X^T X)^{-1}\right)$$

Also
$\vec{E} = \vec{Y} - \hat{\vec{Y}}$
$\Rightarrow \vec{E} = \vec{Y} - X\vec{B}$

the diagonal's $j$th entry

$$\Rightarrow B_j \sim N\left(\beta_j, \sigma^2 (X^T X)^{-1}_{j,j}\right)$$

exact

$$\Rightarrow \frac{B_j - \beta_j}{\sqrt{\sigma^2 (X^T X)^{-1}_{j,j}}} \sim N(0,1)$$ which allows for hypothesis testing and CI's if $\sigma^2$ known.

But what if we don't know $\sigma^2$? This was Student's problem!

Remember how we derived the Student's T-distr?

We used Cochran's thm

Let

$$\vec{Z} \sim N_n(\vec{0}_n, I_n)$$

Let $\vec{\mathcal{E}} := \sigma\vec{Z} = (\sigma I_n) Z \sim N_n\left(\vec{0}_n, (\sigma I_n) I_n (\sigma I_n)^T\right) = N_n(\vec{0}_n, \sigma^2 I_n)$

$$\Rightarrow \vec{Z} = \frac{1}{\sigma}\vec{\mathcal{E}}$$

$$\vec{Z}^T\vec{Z} \sim \chi^2_n, \quad \vec{Z}^T\vec{Z} = \left(\frac{1}{\sigma}\vec{\mathcal{E}}\right)^T\left(\frac{1}{\sigma}\vec{\mathcal{E}}\right) = \frac{1}{\sigma^2}\vec{\mathcal{E}}^T\vec{\mathcal{E}}$$

$$= \frac{1}{\sigma^2}\vec{\mathcal{E}}^T\left((H) + (I-H)\right)\vec{\mathcal{E}} = \frac{1}{\sigma^2}\vec{\mathcal{E}}^T \underbrace{H}_{B_1} \vec{\mathcal{E}} + \frac{1}{\sigma^2}\vec{\mathcal{E}}^T\underbrace{(I-H)}_{B_2}\vec{\mathcal{E}} = \vec{Z}^T H \vec{Z} + \vec{Z}^T(I-H)\vec{Z}$$

We know $B_1 + B_2 = I_n$

$\operatorname{rank}[B_1] = \operatorname{rank}(H) = \dim[\operatorname{colsp}(X)] = p+1$
$\operatorname{rank}[B_2] = \operatorname{rank}[I-H] = \dim[\operatorname{colsp}(X_\perp)] = n-(p+1)$
$\Rightarrow \operatorname{rank}(B_1) + \operatorname{rank}(B_2) = n$

$$\xrightarrow{\text{Cochran's Thm}} \frac{1}{\sigma^2}\vec{\mathcal{E}}^T H \vec{\mathcal{E}} \sim \chi^2_{p+1} \text{ indep. of } \frac{1}{\sigma^2}\vec{\mathcal{E}}^T(I-H)\vec{\mathcal{E}} \sim \chi^2_{n-(p+1)}$$

How does this help?

$$\frac{1}{\sigma^2}\vec{\mathcal{E}}^T(I-H)\vec{\mathcal{E}} = \frac{1}{\sigma^2}\vec{\mathcal{E}}^T(I-H)(I-H)\vec{\mathcal{E}} = \frac{1}{\sigma^2}\vec{\mathcal{E}}^T(I-H)^T(I-H)\vec{\mathcal{E}}$$

$$= \frac{1}{\sigma^2}\left((I-H)\vec{\mathcal{E}}\right)^T(I-H)\vec{\mathcal{E}} = \frac{1}{\sigma^2}\|(I-H)\vec{\mathcal{E}}\|^2 = \frac{1}{\sigma^2}\|(I-H)(\vec{Y} - X\vec{\beta})\|^2$$

$$= \frac{1}{\sigma^2}\|\underbrace{(I-H)\vec{Y}}_{\vec{Y}-\hat{\vec{Y}}} - \underbrace{(I-H)X}_{=\vec{0}_n!}\vec{\beta}\|^2 = \frac{1}{\sigma^2}\|\vec{E}\|^2 \sim \chi^2_{n-(p+1)}$$

$$\Rightarrow E\left[\frac{1}{\sigma^2}\|\vec{E}\|^2\right] = n-(p+1) \quad \text{expectation of } \chi^2_k \text{ r.v.}$$

$$\Rightarrow E\left[\frac{\|\vec{E}\|^2}{n-(p+1)}\right] = \sigma^2$$

$\Rightarrow$ If $MSE := \frac{SSE}{n-(p+1)}$, then MSE is an unbiased estimator for $\sigma^2$.

Let $RMSE = \sqrt{MSE}$